Strategic content production and link formation in socio-computational systems

Yu Zhang and Mihaela van der Schaar
Department of Electrical Engineering, UCLA

Socio-computational Systems – Emergence

– *Social computing* allows individuals to share content, contribute expertise, collectively solve tasks, disseminate information at a low cost.

  • *Knowledge gathering*: IBM Smarter Planet
  • *Knowledge sharing*: Wikipedia
  • *Crowdsourcing*: Amazon Mechanical Turk
  • *Content sharing*: Peer-to-Peer and YouTube
Challenges

Understand and influence how strategic agents proactively make decisions on:

- **Information production** - agents decide whether to personally create and share information
  - “Information” - any knowledge, data, content, service etc.

- **Link formation** - agents decide whether to exchange (provide and receive) information with other agents, i.e. whether to “form links”
Related works - Link Formation

CS literature

  - network formation and price of anarchy in networks with indirect information transmission (agents can access not only information from “neighbors”, but also from neighbors of neighbors)

  - bilateral network formation and price of anarchy in networks where link creation requires mutual consent and cost is two-sided
Related works - Link Formation

**CS literature**

  - proposes efficient (polynomial time) algorithms to find Nash equilibria that are near-optimal given that agents have specific connectivity requirements

  - network formation game with contagious risk, where an agent is exposed to the risk of being hit by a cascading failure based on its connectivity
Related works - Link Formation

CS literature - Limitations

- Assumes agents are exogenously endowed with information and focus solely on strategic link formation, *i.e. the capability of agents to self-produce information & the connection between strategic information production and link formation are neglected.*
Related works - Link Formation

Game-theoretic/Economics literature

- Network games: many interesting works (Jackson, Goyal, etc.)
- Homogeneous agents (Bala and Goyal, 2000)
  - Equilibrium topologies are symmetric: circles, stars, variants of stars.
- Heterogeneous agents (Galeotti and Goyal, 2006)
  - A strict equilibrium is a minimal network, and every minimal network could be a strict equilibrium for some benefits and costs.
- Indirect information flow (Hojman and Szeidl, 2005)
  - When value of information is decaying over the distance from which is acquired, the equilibrium topologies usually have small diameters.

A network topology is defined as the longest path between every pair of agents in the topology.
The Law of the Few (Galeotti and Goyal, 2010)

- It considers strategic decisions of agents on both information production and link formation.
- It predicts that any resulting equilibrium exhibits “the law of the few”
  - As the population of a network grows, the number of information producers is upper-bounded.
  - The network will be dominated by agents who do not personally produce information but only form links and receive information from others.
- Limitations
  - Information is perfectly substitutable.
  - Results are idealized and unrealistic:
    - Most agents do not produce any information (when the network size goes to infinity)
    - The total amount of information produced is a constant, which is invariant to the network size.
The Law of the Few (Galeotti and Goyal, 2010)

- It considers strategic decisions of agents on both information production and link formation.
- It predicts that any resulting equilibrium exhibits “the law of the few”
  - As the population of a network grows, the number of information producers is upper-bounded.
  - The network will be dominated by agents who do not personally produce information but only form links and receive information from others.
Limitations

- Most research in Game Theory and CS assume that agents are endowed with exogenous amounts of information and focus solely on the strategic aspect of link formation.

- Research in Game Theory treat information collected from different agents as equally valuable and perfectly substitutable.

- However, in social computing systems: the agents’ information is endogeneous, heterogeneous, and not perfectly substitutable.
Our Contribution

- A model which captures agents’ strategic behavior on both information production and link formation
- Explicitly consider the information heterogeneity (Dixit-Stiglitz model)
  - An agent’s benefit from information does not only depend on the total amount of information it consumes, but also on the diversity of information it consumes.
- Predict the emergence of core-periphery structures as the network size grows
  - Hub agents produce large amounts of information, have a large number of connections and serve as the major source of information
  - Spoke agents produce and share limited amounts of information and mainly consume information acquired from hub agents
  - Law of the few disappears with information heterogeneity
Model

- We consider a social computing system consisting of $n$ agents, where individual agents can share information they produce with each other.
- $N \triangleq \{1, \ldots, n\}$: set of agents
- $x_i \in \mathbb{R}^+$: agent $i$’s production level
  - The amount of information produced by agent $i$
- $g_i \triangleq g_{i1}, \ldots, g_{in} \in \{0,1\}^n$: link formation decision of agent $i$
  - $g_{ij} = 1$: agent $i$ forms a link to agent $j$
  - $\bar{g}_{ij} = \max g_{ij}, g_{ji} = 1$: agents $i$ and $j$ are connected (neighbors)
- Consider unilateral link formation
  - Links are created by the unilateral actions of agents, and link costs are one-sided
- First: direct information sharing
  - Each agent can access and consume all information produced by its neighbors.
Other Variables

• \( s = x, g \): a strategy profile
  - \( x = x_i^{n} \) : decisions on information production
  - \( g = g_i^{n} \) : decisions on link formation

• \( \bar{g} \triangleq \left[ \bar{g}_{ij} \right]_{i,j \in N} \): the connectivity graph

• \( N_i \ g \triangleq j \mid g_{ij} = 1 \) : the set of agents who agent \( i \) forms links to
  - Determines the total link formation cost of agent \( i \)

• \( N_i \ \bar{g} \triangleq j \mid \bar{g}_{ij} = 1 \) : the set of agents who agent \( i \) connects with
  - Determines the information that agent \( i \) can consume
Utility Function

- An agent’s total utility

\[ u_i(x, g) = v\left(\left( x_i^0 + \sum_{j \in N_i} x_j^0 \overline{g} \right)^{1/\rho}\right) - cx_i - k|N_i - g| \]

the benefit from information consumption

- \( \rho \in 0, 1 \) measures how much information variety matters to agent
  - When \( \rho = 1 \), information is perfectly substitutable and variety does not matter
  - When \( \rho \) is small, variety matters a lot

\[ X_i \triangleq \left[ x_i^0 + \sum_{j \in N_i} x_j^0 \overline{g} \right]^{1/\rho} \]: the amount of agent i’s effective information
Utility Function

• An agent’s total utility

\[ u_i(x, g) = v \left[ x_i^p + \sum_{j \in N_i} x_j^p \right]^{1/\rho} - c x_i - k \left| N_i \right| g \]

- cost for information production
- cost for link formation
Assumptions on Utility Function

• \( v \) is a twice continuously differentiable, increasing, and strictly concave function.
  – The benefit from information consumption saturates when the amount of effective information consumed is sufficiently large.

• \( v \) satisfies
  – \( v(0) = 0 \)
  – \( \lim_{x \to \infty} v'(x) = 0 \)
  – \( v'(0) \triangleq \lim_{x \to 0^+} v(x) < \infty \)
  – \( v'(0) > \alpha \) where \( \alpha > 0 \) is a constant
Assumptions on Utility Function

• \( v(\cdot) \) is a twice continuously differentiable, increasing, and strictly concave function.
  - The benefit from information consumption saturates when the amount of effective information consumed is sufficiently large.

• \( v(\cdot) \) satisfies
  - \( v(0) = 0 \)
  - \( \lim_{x \to \infty} v'(x) = 0 \)
  - \( v'(0) \triangleq \lim_{x \to 0^+} v(x) < \infty \)
  - \( v'(0) > \alpha \) where \( \alpha > 0 \) is a constant

These assumptions:
- an agent’s benefit increases with the amount of effective information, but the rate of increase decreases and approaches 0.

Hence, there always exists an upper bound on the amount of information that an individual agent is willing to acquire given the cost of information production and link formation.
Assumptions on Utility Function

\[ \frac{\partial^2 u_i}{\partial x_i \partial x_j} \bigg|_{x_i, x_j \in \mathbb{R}^+} < 0, \ \forall i \text{ and } \forall j \in N_i \quad \overline{g} \]

Substitutability among agents’ information:
an agent’s marginal benefit of production decreases against the amount of information produced by its neighbors.

Hence, the more information an agent acquires from its neighbors, the less incentive it has to produce information by itself.
Information production & link formation game

- Simultaneous move game where agents make decisions on information production and link formation at the same time.

Equilibrium analysis for the basic model

- We consider pure strategies.
- A strict Nash equilibrium is a strategy profile \( s^* = x^*, g^* \) such that

\[
\begin{align*}
    u_i (s_i^*, s_{-i}^*) &> u_i (s_i, s_{-i}^*), \quad \forall s_i \in \mathbb{R}^+ \times [0,1]^{n-1}, \forall i \in N
\end{align*}
\]

- Lemma 1: Given \( c, k, \rho, \) and \( v, \) in any equilibrium \( s^* = x^*, g^* , \)
  - \( g_{ij}^* g_{ji}^* = 0 \) for all \( i, j \in N \)
  - \( x_i^* > 0 \) for all \( i \in N \)
  - \( x_i^* \leq \overline{x} \) for all \( i \in N, \) where \( \overline{x} \) is the unique solution of \( v' \overline{x} = c \)
Information production & link formation game

An agent always produces a positive amount of information, since the information it acquires from neighbors can never fully replace its own information.

An agent’s production is upper bounded \( \Rightarrow \) the maximum equilibrium production

**Lemma 1:** Given \( c, k, \rho, \) and \( v, \) in any equilibrium \( s^* = x^*, g^*, \)
- \( g_{ij}^* g_{ji}^* = 0 \) for all \( i, j \in N \)
- \( x_i^* > 0 \) for all \( i \in N \)
- \( x_i^* \leq \bar{x} \) for all \( i \in N, \) where \( \bar{x} \) is the unique solution of \( v' \bar{x} = c \)
Equilibrium analysis for the basic model

Proposition 1. Given $c$, $k$, $\rho$, and $v$, each equilibrium belongs to one of the following two types:

(i) each agent personally produces an amount $\bar{x}$ of information and no agent forms any links;

(ii) each agent personally produces an amount strictly smaller than $\bar{x}$ and is connected with at least one other agent.

Focus on non-trivial case – non-empty network

Next: separate agents into 2 types depending on production levels to characterize interactions between high and low producers at equilibrium.
Separate agents depending on production

- For a strategy profile, we order agents by their production levels \( x_1 \geq x_2 \geq \cdots \geq x_n \)

- Two types of producers:
  - A positive integer \( n_h \) such that \( n \) exists such that
    - Highest producers: \( x_i = x_{n_h} \) for all \( i \leq n_h \)
    - Low producers: \( x_j < x_{n_h} \) for all \( j > n_h \)

- We assume that \( n_h < n \)
  - A strategy profile is a symmetric profile if \( n_h = n \)
  - We also analyzed symmetric profiles, but this is omitted here.
Properties at equilibrium: High producers

- Lemma 2: In an equilibrium $s^* = x^*, g^*$, the following properties hold for $g^*$:
  - $\bar{g}_{ij}^* = 0$, for some $i, j \in N$ and $i \neq j$
    - The network is never complete at equilibrium
    - If the network was complete, the profile would be symmetric
  - For each $i \leq n_h s^*$, $g_{ij}^* = 0$, $\forall j > n_h s^*$
    - No high producer forms links to low producers
  - For each $i \leq n_h s^*$, $\bar{g}_{ii'}^* = 0$, for some $i' \leq n_h s^*$ and $i' \neq i$
    - High producers are not all linked to each other
Properties at equilibrium: Low producers

- Lemma 3: In an equilibrium $s^* = x^*, g^*$, the following properties hold for $g^*$:
  - For each $j > n_h$ $s^*$, $g^*_{ji} = 1$, for some $i \leq n_h$ $s^*$.
  - Each low producer forms links to at least one high producer $g^*_{jj'} = 1$, for some $j, j' > n_h$ $s^*$ only if $g^*_{ji} = 1$, $\forall i \leq n_h$ $s^*$.
  - A low producer will only form a link to another low producer if it is already connected to all high producers.
Further classification

We can classify low producers into 2 types:

**Type 1:** A low producer forming links to all high producers and to some of the low producers;

**Type 2:** A low producer forming links with some (or all) high producers and no low producer.

Next: we prove that when network size is sufficiently large, low producers of Type 1 will disappear in any equilibrium. Hence, only high producers and low producers of Type 2 exist at equilibrium.
Equilibrium Topology

- Theorem 1: Given $c$, $k$, $\rho$, and $\nu$, and when the network size $n$ goes to infinity
  - Only two types of agents exist in any equilibrium: a hub agent is a high producer who produces an amount $\tilde{x}x^*$; a spoke agent is a low producer who produces an amount $\tilde{x}x^*$ and forms links with $qg^*$ hub agents;

$$
\lim \inf_{n \to \infty} n_h^* \inf_{s^* \in S_n^*} n_h^* = \infty \quad \text{and} \quad \lim \inf_{n \to \infty} n_l^* \inf_{s^* \in S_n^*} n_l^* = \infty \quad \text{where} \quad n_l^* s^* \triangleq n - n_h^* s^*
$$

and $S_n^*$ denote the set of equilibrium strategy profiles when the network size is $n$.

- Key idea:
  - The maximum number of agents that a high producer can support is upper-bounded.

The law of the few disappears!
Equilibrium Topology

- Theorem 1: Given $c$, $k$, $\rho$, and $\nu$, and when the network size $n$ goes to infinity
  - Only two types of agents exist in any equilibrium: a hub agent is a high producer who produces an amount $\tilde{x} x^*$; a spoke agent is a low producer who produces an amount $x x^*$ and forms links with $q g^*$ hub agents;

The law of the few disappears!
Equilibrium Topology (cont’d)

- Theorem 2: Given $c$, $k$, $\rho$, and $\nu$, the number of hub agents grows at the same order as the entire population, i.e. \( \inf_{s^* \in S_n^*} n_{h^*} \) is \( \Omega n \)
  - There are two constants $\delta_1$ and $\delta_2$ such that
    \[
    \delta_1 n \leq \inf_{s^* \in S_n^*} n_{h^*} \leq \delta_2 n, \ \forall n
    \]
Illustrative results

When $n$ is sufficiently large, the number of spoke agents that each hub agent can support reaches its upper bound, and the fraction stops decreasing.

A spoke agent mainly relies on information acquired from others rather than self-production and will be more significantly influenced by the change on the network size.
The impact of $\rho$

High preference for diversity $\rightarrow$ Low preference for diversity

Average degree

$k = 0.1$
$k = 0.3$
$k = 0.5$
Equilibrium Topology (cont’d)

• Theorem 3: Given $c$, $k$, $\rho$, and $\nu$, the total amount of information produced in the network at equilibrium, i.e. $\sum_{i \in N} x_i^*$, grows at the same order as the population size $n$. That is

$$\inf_{s^* \in S_n^*} \left\{ \sum_{i \in N} x_i^* \right\} \text{ and } \sup_{s^* \in S_n^*} \left\{ \sum_{i \in N} x_i^* \right\} \text{ are } \Omega \ n$$

The law of the few disappears!
Illustrative results –
total amount of information produced at equilibria

Law of the few
Social Optimum

• The social welfare of a strategy profile

\[ W(x, g) = \sum_{i \in N} u_i(x, g) \]

• Lemma 3: Given \( c, k, \rho, \) and \( v, \) the social optimum is upper-bounded as follows

\[ W^\# \leq \max \left[ n \left[ v n^{1/\rho} \hat{x}_n - c \hat{x}_n \right] - nk / 2, n \left[ v \bar{x} - c \bar{x} \right] \right] \]

– \( \hat{x}_n \) is the solution of \( v' n^{1/\rho} \hat{x}_n = c / n^{1-\rho}/\rho \)

the upper bound when the optimal network is empty

the upper bound when the optimal network is non-empty
Price of Stability

- Theorem 4: Given $c$, $k$, $\rho$, and $v$, the Price of Stability is upper-bounded as

$$PoS \leq \max \left[ v \frac{1}{\rho} \bar{x}_n - cx_n \right] - \frac{k}{2} / \left[ v \bar{x} - c\bar{x} \right], 1$$

- Idea of proof:
  - The social welfare $W_{x^*, g^*}$ of any equilibrium strategy is lower bounded by $n \left[ v \bar{x} - c\bar{x} \right]$

- Corollary 1: The possible loss (POS upper bound) monotonically increases with $n$

- This is because the set of available strategies increases with $n$, whereas the set of eq. strategies remains limited as each hub in an eq. will only support a limited number of spokes.
Can we improve the PoS?

- The cost of forming a link can be adjusted and more efficient equilibrium can be induced -> pricing on links

<table>
<thead>
<tr>
<th>Normalized social welfare</th>
<th>ρ</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-5 -4 -3 -2 -1 0 1 2 3 4 5 0.5 1.0</td>
</tr>
</tbody>
</table>

- subsidize the link creator to increase agents’ incentive to form links and decrease self-production
- charge the link creator to reduce agents’ incentive to form links and increase self-production
Indirect Information Sharing

• We now consider the case where an agent can also consume information which its neighbors acquired from other agents.

• Agent utility

\[
\begin{align*}
    w_i \cdot x, g & = v \left[ x_i^\rho + \left( \sum_{l=1}^{n-1} \sum_{j \in N_i} x_j^\rho \right)^{1/\rho} \right] - cx_i - k \left| N_i \cdot g \right|
\end{align*}
\]

• Theorem 5: Given \( c, k, \rho, \) and \( v \), there always exits at least one equilibrium in the presence of indirect information sharing
Equilibrium

- Theorem 6: There is a value $\bar{k}_n$ such that
  
  - When $k > \bar{k}_n$, there exists a unique equilibrium where each agent personally acquires an amount $\bar{x}$ of information and no agent forms links
  
  - When $k < \bar{k}_n$, then each equilibrium is minimally connected
    - Minimally connected network: There is a unique path between every pair of agents on its connectivity graph

Key difference to direct information sharing:
here the agents with most connections become the network’s hubs and share the information unlike in the direct information sharing (high producers are the hubs)
Conclusions

• We have studied agent behavior in social computing systems.

• If information diversity matters, and information production and link formation are choices:
  – Many hubs and many spokes
  – Law of the few disappears
  – Total information production grows as the same rate as the size of the network

• Strategic design – can improve performance

• Information dissemination – new results

• Economics implications?